remote sensing

MDPI

Article

Predicting carbon accumulation in temperate forests of Ontario, Canada using a LiDAR-initialized growth-and-yield model, supplementary material

Paulina T. Marczak [1]*, Karin Y. van Ewijk [1], Paul M. Treitz [1], Neal A. Scott [1], and Donald C.E. Robinson [2]

[1] Department of Geography and Planning, Queen's University, Kingston K7L 3N6, Canada; karin.vanewijk@queensu.ca (K.Y.V.E.); paul.treitz@queensu.ca (P.M.T.); neal.scott@queensu.ca (N.A.S.)
[2] ESSA Technologies; drobinson@essa.com (D.C.E.R.)
* Correspondence: paulina.marczak@queensu.ca (P.T.M.)

Received: 02 December 2019; Accepted: 03 January 2020; Published:

S1: Deriving basal area/diameter conversion formula

To further explain the logic behind the relationship of BA and DBH, the formula presented in the main text (equation 1) is shown below:

$$\text{Basal area} = ((\pi DBH^2/40{,}000))/0.0625$$

The derivation of this formula follows below.

Because BA is a measure of circular area, diameter should first be converted using the formula for the area of a circle:

$$\text{Area of a circle} = \pi(\text{radius})^2$$

To convert from diameter to area, some modification is needed. As diameter is twice the radius, the above can be rewritten as:

$$\text{Area of a circle} = (\pi(\text{diameter}/2)^2$$

This can be simplified to:

$$\text{Area of a circle} = \pi \text{diameter}^2/4$$

However, it is important to note that DBH is reported and measured in centimeters, BA is typically in m^2, or 10,000 cm^2 per m^2. The formula must account for length to area conversion by multiplying the length units by 10,000. Thus,

$$\text{Basal area} = (\pi DBH^2/40{,}000)$$

BA is expressed as m^2/ha, thus the values were divided by the plot area (625m^2, or 0.0625 ha) to obtain the average plot-level measurement, giving:

$$\text{Basal area} = ((\pi DBH^2)/40{,}000))/0.0625$$

The above was modified from conversation with Karin van Ewijk (personal communication, April 6, 2018), [1], and [2].

Similarly, average DBH could be derived from BA_dist and SDD or SD by dividing BA_dist by SD or SDD, giving average individual-tree BA. Then, average DBH can be found by rearranging the formula for basal area above (i.e., equation 2,3 in main text).

S2: Measures of evaluation descriptions

R^2 was calculated from cross-validation as:

$$R^2 = 1 - \frac{SSE}{SST}$$

where SSE, or Sum of Squared Errors, is the sum of the squared differences between observed and predicted values (i.e., explained variation). SST is Total Sum of Squares, or the predicted value subtracted from the average actual value (i.e., total variation).

Relative RMSE (rRMSE) was used a measure of model performance across varying size classes, calculated by dividing the RMSE by the true mean of the observed values [3,4], and was written as:

$$rRMSE = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}} \; / \; \bar{y}$$

where y is the observed value, ŷ is the predicted value, ȳ is the mean of observed values, and *n* is the number of observations.

Scaled Root Mean Square Distance (sRMSD)[5] was used to compare accuracy across size classes. It is interpreted as RMSE scaled by the standard deviation of observed values, producing a value between 0 and 1:

$$sRMSD = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n}} \; / \; \sigma_y$$

where *y* is the observed value, *ŷ* is the predicted value, *n* is the number of observations, and σ_y is the standard deviation of observed values. Shang et al. (2017) [3] give reasoning for sRMSD, in that predictive accuracy can be measured against the overall dispersion (i.e. standard deviation) of the response, rather than the mean (i.e., rRMSE) [3]. rRMSE would provide optimistic measures of accuracy when the variance of the data is much smaller than the mean.

BEI is written as such:

$$BEI = \left(\frac{\sum_{i=1}^{k}|f_i - \hat{f}_i|}{n} + \frac{\sum_{i=1}^{k}|f_i - \hat{f}_i|}{\hat{n}}\right) \Big/ 2$$

where f_i is the observed number of stems in class *i*, *k* is the total number of classes, and $\hat{f}$ is the predicted number of stems in class *i*. *n* represents the observed total number of stems across all size classes, and $\hat{n}$ represents the total predicted number of stems across all size classes. BEI was developed to address issues with previous iterations of an error index for SDD, as it effectively limits the index between 0 and 1, does not penalize based on incorrect prediction of stem density, and is not reliant on wood volume, stem density, or dollar value [3].

Bias was calculated as:

$$\%Bias = \left(\frac{1}{n}\sum\frac{y-\hat{y}}{\hat{y}} \times 100\right)$$

Where where *y* is the observed value, *ŷ* is the predicted value, and *n* is the number of observations.

S3: Walk-through of methodological approach using subset of input data for PlotID 'PRF003'- Sapling Class- FVS2 model, constant mortality

To provide additional guidance on the G&Y model parameterization and execution, a brief walk-through using sample data is shown below. This walk-through starts after LiDAR-predicted variables have already been generated and average DBH derived. It also has species abundance data already applied to the generated average DBH and the data has already been processed into .xls files generated for each individual plot, i.e., the tree lists, using an automated script and the writeXLS library in R. Named ranges for each file (i.e., each plot for each model) were added manually in order for the TLM to read them; we could not find an automated way to do this. For more information and detail on FVS model parameterization, the reader is referred to Woods and Robinson 2007 [6].

Firstly, we select the pre-processed treelist_PRF003_sapling_FVS2.xls, sample data shown below:

Plot	DBH	Species	TREEID	STATUS
PRF003	11.92156	BF	1	L
PRF003	11.92156	BF	2	L
PRF003	11.92156	BF	3	L
PRF003	11.92156	BF	4	L
PRF003	11.92156	BF	5	L
PRF003	11.92156	BF	6	L
PRF003	11.92156	MR	7	L
PRF003	11.92156	PW	8	L
PRF003	11.92156	SW	9	L
PRF003	11.92156	SW	10	L
PRF003	11.92156	SW	11	L
PRF003	11.92156	SW	12	L
PRF003	11.92156	SW	13	L
PRF003	11.92156	SW	14	L
PRF003	11.92156	SW	15	L
PRF003	11.92156	SW	16	L
PRF003	11.92156	SW	17	L
PRF003	11.92156	PT	18	L
PRF003	11.92156	PT	19	L
PRF003	11.92156	PT	20	L
PRF003	11.92156	PT	21	L
PRF003	11.92156	PT	22	L
PRF003	11.92156	PT	23	L
PRF003	11.92156	PT	24	L
PRF003	11.92156	PT	25	L
PRF003	11.92156	PT	26	L
PRF003	11.92156	PT	27	L
PRF003	11.92156	PT	28	L
PRF003	11.92156	PT	29	L
PRF003	11.92156	PT	30	L
PRF003	11.92156	PT	31	L
PRF003	11.92156	PT	32	L
PRF003	11.92156	PT	33	L
PRF003	11.92156	PT	34	L
PRF003	11.92156	PT	35	L
PRF003	11.92156	PT	36	L
PRF003	11.92156	BW	37	L

Figure S3-1. A sample data tree list generated from LiDAR average DBH and inventory species abundance (FVS2) for saplings in Plot ID PRF003, where status represents whether the tree is living or dead.

We then input the tree list into a second pre-processing step, the Tree List Manager (TLM) extension of $FVS^{Ontario}$, which converts the data to a machine-readable format. As $FVS^{Ontario}$ requires a specific file extension to generate estimates of carbon stock and accumulation, input tree lists were converted from .xls to the .tre extension using the Tree List Manager (TLM) extension provided with $FVS^{Ontario}$. To initialize Project setup, Plot Type was set to Fixed Area Plot. Species code was set to Alpha, and FVS Setup was set to Export Data to FVS. Plot info data included an input list of all plot names and plot area (i.e. 625 m^2). Following this, the corresponding plot would be imported into the TLM Import window. The file was then exported manually with the corresponding plot name as a prefix.

Headers are not included in model parameterization. Left to right, the columns represented in Figure S3-2 are the Plot ID, Tree ID, Tree Status (1=Live, 8 = Dead), Species code, DBH, Top Height (0 is ignored), the Status (left blank to represent the default 'Live' value) , the Damage code (5500 is ignored), and the cutting code (0 is ignored).

```
0001  116.00 1BF 11.9     0.0              5500        0
0001  216.00 1BF 11.9     0.0              5500        0
0001  316.00 1BF 11.9     0.0              5500        0
0001  416.00 1BF 11.9     0.0              5500        0
0001  516.00 1BF 11.9     0.0              5500        0
0001  616.00 1BF 11.9     0.0              5500        0
0001  716.00 1MR 11.9     0.0              5500        0
0001  816.00 1PW 11.9     0.0              5500        0
0001  916.00 1SW 11.9     0.0              5500        0
0001 1016.00 1SW 11.9     0.0              5500        0
0001 1116.00 1SW 11.9     0.0              5500        0
0001 1216.00 1SW 11.9     0.0              5500        0
0001 1316.00 1SW 11.9     0.0              5500        0
0001 1416.00 1SW 11.9     0.0              5500        0
0001 1516.00 1SW 11.9     0.0              5500        0
0001 1616.00 1SW 11.9     0.0              5500        0
0001 1716.00 1SW 11.9     0.0              5500        0
0001 1816.00 1PT 11.9     0.0              5500        0
0001 1916.00 1PT 11.9     0.0              5500        0
0001 2016.00 1PT 11.9     0.0              5500        0
0001 2116.00 1PT 11.9     0.0              5500        0
0001 2216.00 1PT 11.9     0.0              5500        0
0001 2316.00 1PT 11.9     0.0              5500        0
0001 2416.00 1PT 11.9     0.0              5500        0
0001 2516.00 1PT 11.9     0.0              5500        0
0001 2616.00 1PT 11.9     0.0              5500        0
0001 2716.00 1PT 11.9     0.0              5500        0
0001 2816.00 1PT 11.9     0.0              5500        0
0001 2916.00 1PT 11.9     0.0              5500        0
0001 3016.00 1PT 11.9     0.0              5500        0
0001 3116.00 1PT 11.9     0.0              5500        0
0001 3216.00 1PT 11.9     0.0              5500        0
0001 3316.00 1PT 11.9     0.0              5500        0
0001 3416.00 1PT 11.9     0.0              5500        0
0001 3516.00 1PT 11.9     0.0              5500        0
0001 3616.00 1PT 11.9     0.0              5500        0
0001 3716.00 1BW 11.9     0.0              5500        0
```

Figure S3-2: A sample data tree list input into FVS^{Ontario}, list generated from LiDAR average DBH and inventory species abundance (FVS2) for saplings in Plot ID PRF003, saved with the extension `treelist_PRF003_sapling_FVS2`.tre.

The file was then input into FVS$^{\text{Ontario}}$ Fire and Fuels (FFE) Carbon Submodel extension using a batch process and a '.key' file generated from the FVS$^{\text{Ontario}}$ Silviculture Prognosis Interface. The parameters used for constant mortality is as follows in Figure S3-3, noting the number of spaces represents different parameters:

SCREEN

TREELIST 0 1

BAMAX 999

***SDIMAX 900**

***MORTMULT 1 5.9**

FIXMORT 0 0.004

INVYEAR 2012

TIMEINT 0 1

NUMCYCLE 10

Figure S3-3: Part of 'carbon.key' parameterization file used for generating projections of carbon stock.

A Python script [7] was used to automate each plot or size-class level prediction of carbon accumulation. As the script requires the Linux processing environment to run, a virtual machine using Ubuntu v18.04.1 x amd64 disk image was installed, with an allocated memory of 1024 MB and 10 GB virtual hard disk image on a dynamically allocated storage environment. To run the x64 bit version of Ubuntu, the Virtualization option on the host system BIOS settings were changed to "Enabled", while the Windows Feature "Hyper-V" was disabled.

To run FVS$^{\text{Ontario}}$ (i.e., a Windows-native program), Wine v3.0.4 was installed on the virtual machine using the installation procedure for versions prior to Ubuntu v18.10 [8]. Xdotool was installed to simulate input typing into FVS$^{\text{Ontario}}$ [9]. For a given model (e.g.; FVS2 Saplings Constant Mortality), folders with all input tree list (.tre) files ("in") and an output folder for the predictions ("out") were created. In this case, all sapling tree lists, including our Plot ID PRF003, for the 75 plots generated for model FVS2 would be in this folder. All scripts were placed in the same directory as the FVS$^{\text{Ontario}}$ executable. A terminal was opened to set up the automation process:

```
While true
Do
Wine FVSOntario.exe
Done
```

A second terminal was opened in the same directory to run the command "python3 gen_xdotool.py" to generate the xdotool script. The script was then run using the command "bash xdo.sh" to generate the output files. With a ten-second lag in place, the first terminal was then reopened so the process could be run. A detailed description of the automation process can be found in McVittie (2018) [7]. The outputs were unstructured text files, separated per plot with various file and growth information, including values of total aboveground carbon stocks for each year of the projection.

To extract values of carbon stocks, a section of the output text file had to be extracted, compounded by the issue that the file did not have consistent heading structure and the carbon stock values were halfway within the unstructured file. The extraction was automated using the stringr library as well as the grep function in R.

Thus, an output file (.csv) was generated with a column representing Plot ID, a column representing the year of the projection, and a third column with carbon stock values. A separate file was generated for each model iteration, for each mortality rate, and for each size class, including the plot-level average and aggregate predictions. These data were aggregated per mortality rate, size class, and model number for analysis, shown in Figure S3-4.

UniqueID	Year	Carbon	PlotID
10	2021	9.9	prf003
101	2021	6	prf007
102	2021	7.7	prf009
103	2021	9.2	prf024
104	2021	8.7	prf026
105	2021	15.9	prf027
106	2021	7.9	prf028
107	2021	9.9	prf030
108	2021	12.8	prf031
109	2021	7.5	prf033
1010	2021	8.7	prf036
1011	2021	4.7	prf037
1012	2021	10.3	prf038
1013	2021	5.9	prf039
1014	2021	11.8	prf040

Figure S3-4: Carbon stock outputs (tons/ha) after pre-processing FVS^{Ontario} raw file outputs, where saplings are represented by uniqueID 10 for PlotID PRF003 for FVS2.

To compare model outputs with each other, we ran the Robinson test on the outputs (such as shown in Figure S3-4) comparatively using the equivalence library (equiv.boot function) in R. For our sapling example, we would also seed our model to obtain consistent results:

```
Set.seed(123)
```

Next, the Robinson test of equivalence was performed for both carbon stock and accumulation. The following is based on a newly created table where each column represents values of carbon stock or accumulation for a particular year for a given model. The "`FVS1_2012_Sapling_Carbon`" column represents the validation data, and the "FVS2_2012_Sapling_Carbon" column represents the 'carbon' column in Figure S3-4.

```
Robinson_test = equiv.boot(table$FVS1_2012_Sapling_Carbon,table$FVS2_2012_Sapling_Carbon)
```

For carbon accumulation, the resulting stock values at year 1 would be subtracted from year 2 for statistical equivalence testing.

The additional graphical outputs demonstrating equivalence of bias and proportionality were generated using scripts from Fekety (2019)[10] and may become available upon request.

S4: Additional table for Figure 11

Table S4. Measures of accuracy and precision for FVS3 carbon stocks across plot-level models compared to FVS2-Average carbon stocks at year 2012. 2016, and 2021 carbon stock values appear in brackets, with constant mortality (All forest types). Table represents statistical descriptors for Figure 11 in main text. Mean of observed values are 30.40, 33.48, and 37.45 tons/ha, respectively.

Model	RMSE	rRMSE	sRMSD	R	Bias (%)	Mean
FVS2 Aggregate	13.95(13.68;13.32)	0.46(0.41;0.36)	1.22(1.18;1.11)	0.64(0.64;0.65)	5.15(3.47; 0.58)	31.97 (34.64;37.67)
FVS3 Average	2.98(3.28;3.76)	0.10(0.10;0.10)	0.26(0.28;0.31)	0.97(0.96;0.95)	0.37(0.75;1.07)	30.52(33.74; 37.85)

FVS3 Aggregate	14.91(14.56;14.16)	0.49(0.43; 0.38)	1.31(1.26;1.18)	0.64(0.65;0.66)	8.21(6.64;3.81)	32.90(35.71; 38.88)

References

1. Erdle, T. n.d.. *Measurement of Tree Basal Area and Volume* [PowerPoint presentation]. Retrieved from http://ifmlab.for.unb.ca/People/Kershaw/Courses/For1001/Erdle_Version/TAE-BasalArea&Volume.pdf
2. Lim, K.; Treitz, P.; Baldwin, K.; Morrison, I.; Green, J. Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing* **2003**, *29*, 658-678.
3. Shang, C.; Treitz, P.; Caspersen, J.; Jones, T. Estimating stem diameter distributions in a management context for a tolerant hardwood forest using ALS height and intensity data. *Can. J. Remote Sens*. **2017,** *43*, 79-94.
4. Packalén, P.; Maltamo, M. Estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Can. J. For. Res.* **2008,** *38*, 1750-1760.
5. Stage, A. R.; Crookston, N. L. Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. *Forest Science* **2007**, *53*, 62-72.
6. Woods, M.; Robinson, D. Development of FVSOntario: a Forest Vegetation Simulator variant and application software for Ontario. In *USDA Forest Service Proceedings RMRS-P-54. 2008, Third Forest Vegetation Simulator Conference*, Fort Collins, CO, USA, 2007 February 13–15; Havis, R.N.; Crookston, N.L.; US Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2008.
7. McVittie, A. *FVSAutomator*. Available online: https://github.com/mcvittal/FVSAutomator (accessed on 10 January 2019).
8. Wine HQ [Computer software]. **2018**. Retrieved from https://wiki.winehq.org/Ubuntu
9. Sissel, J. **2009.** Xdotool [Computer software]. Retrieved from https://www.semicomplete.com/projects/xdotool/
10. Fekety, P. *Procedures for obtaining graphical outputs of equiv.boot from equivalence package*; Version 1.0. [R code].